Monte Carlo Sampling Methods Homework 1

Steven (Zihan) Zhang

September 21, 2022

I have discussed and consulted with Qianyu Zhu and Linkai Ma while doing this homework. I also read Larry Wasserman's book *All of Statistics* for some reference.

0.1 Part 1

When N becomes bigger in magnitude, the distribution of $\sqrt{N}(\bar{x}_n - \pi[x])$ becomes more normally distributed (symmetric and centralized to 0). Also, the QQ plot would become more linear, meaning it gradually converges to $N(0,1)$ in distribution. Figure 1 is the graph for illustration.

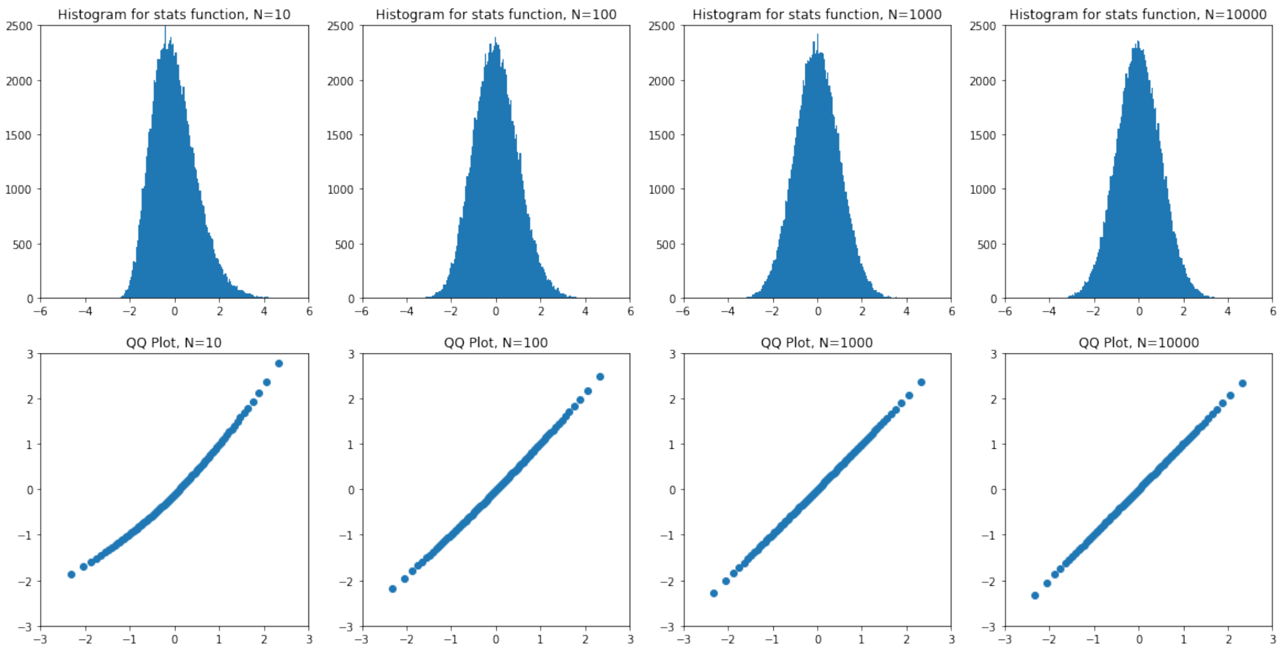

Figure 1: Changes in empirical distribution of $\sqrt{N}(\bar{x}_N - \pi[x])$ when N grows in magnitude from 10^1 to 10^4 (top) and the corresponding QQ plot of target statistics against normal distribution (bottom).

0.2 Part 2

Suppose $Y \sim \Gamma(N, \frac{1}{N})$, we could define:

$$\bar{x}_N = \frac{1}{N}\sum_{i=1}^{N} X^{(i)} \sim \Gamma(N, \frac{1}{N}) \tag{1}$$

$X^{(i)}$ is from exponential distribution, $X^{(i)} \sim \exp(1)$. We then formulate the distribution:

$$p_N = \mathbb{P}[\bar{x}_N - \pi[x] > 0.1] = \mathbb{P}[\bar{x}_N > 1.1] = \mathbb{P}[\bar{x}_N > 1 + \epsilon] \tag{2}$$

where $\epsilon = 0.1$. Thus,

$$\gamma = \lim_{N\to\infty} \frac{1}{N} \log \mathbb{P}(\bar{x}_N - \pi[x] > \epsilon) = -\epsilon + \log(1+\epsilon) \tag{3}$$

which is approximately 0. Thus, a suitable unbiased estimator Q_N for the actual activity is:

$$Q_N = \frac{1}{M}\sum_{i=1}^{N} 1_{\{Y^{(i)} > 1.1\}} = \frac{1}{M}\sum_{i=1}^{N} 1_{[1.1,\infty]}(Y^{(i)}) \tag{4}$$

where $Y^{(i)} \sim \Gamma(N, \frac{1}{N})$ are independent and identically distributed. M is a fixed sample number, which equals 10^8 in our calculation. We also set the range of N as $N < 2^{14}$. Using that estimator Q_N for activity q_N, we compared the estimated exponential decay rate of p_N and its exact value given by internal *cdf* function in *scipy* in Figure 2.

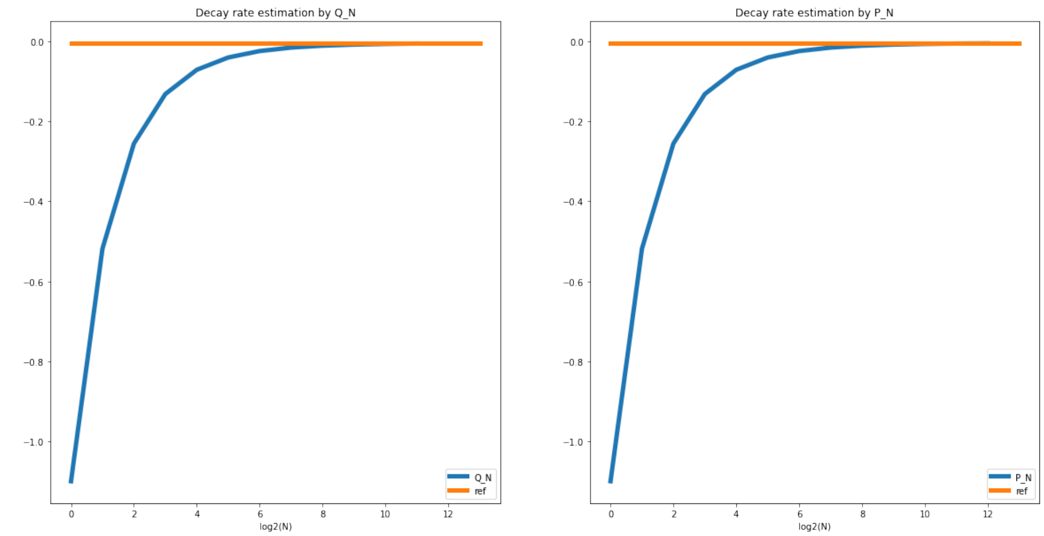

Figure 2: Comparison of of estimated exponential rate of p_N using Q_N (left) and the same rate using p_N given by *scipy cdf* function.

We also calculate the standard deviation of Q_N using p_N:

$$\begin{aligned}\sigma(Q_N) &= \sqrt{\text{Var}\left(\frac{1}{M}\sum_{i=1}^{M} 1_{\{(\bar{x}_N)_i > 1.1\}}\right)} \\ &= \sqrt{\frac{\text{Var}(1_{\{(\bar{x}_N)_i > 1.1\}})}{M}} \\ &= \sqrt{\frac{p_N(1-p_N)}{M}}\end{aligned} \tag{5}$$

When $p_N \to 0$:

$$\frac{\sigma(Q_N)}{p_N} \approx \frac{1}{\sqrt{M \cdot p_N}} \tag{6}$$

Thus, to make sure the solution is not dominated by error, we should select $M \sim 1/p_N$ as a fixed constant, or run the simulation with small N. Figure 3 illustrates the comparison of decay rates.

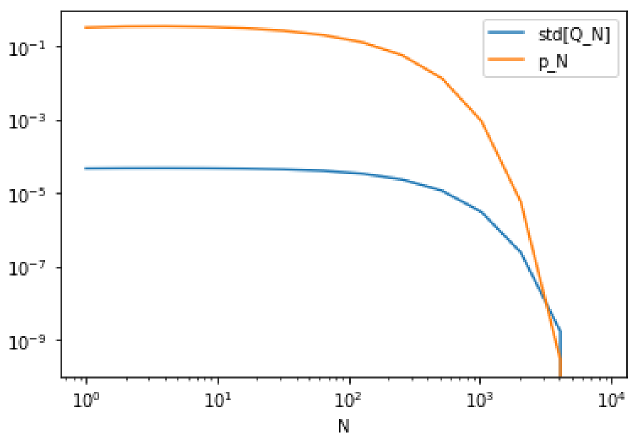

Figure 3: Comparison of decay rate of Q_N and p_N under different N at fixed M. The standard deviation of Q_N would decay much slower.